



Syddansk Universitet

Sharing data increases citations

Drachen, Thea Marie; Ellegaard, Ole; Larsen, Asger Væring; Dorch, Bertil F.

Published in:

LIBER Quarterly - Journal of European Research Libraries

Publication date:

2016

Document version

Final published version

Document license

CC BY-NC-ND

Citation for pulished version (APA):

Drachen, T. M., Ellegaard, O., Larsen, A. V., & Dorch, B. F. (2016). Sharing data increases citations. LIBER Quarterly - Journal of European Research Libraries, 26(2), 67-82.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Sharing Data Increases Citations

Thea Marie Drachen

University of Southern Denmark, Odense
thmd@bib.sdu.dk

Ole Ellegaard

University Library of Southern Denmark, Odense
oleell@bib.sdu.dk

Asger Væring Larsen

University Library of Southern Denmark, Odense
avla@bib.sdu.dk

Søren Bertil Fabricius Dorch

University Library of Southern Denmark, Odense
bfd@bib.sdu.dk

Abstract

This paper presents some indications to the existence of a citation advantage related to sharing data using astrophysics as a case. Through bibliometric analyses we find a citation advantage for astrophysical papers in core journals. The advantage arises as indexed papers are associated with data

by bibliographical links, and consists of papers receiving on average significantly more citations per paper per year, than do papers not associated with links to data.

Key Words: data; data sharing; citation; linking

1. Introduction

Proper data management including sharing of data is vitally connected to the ethics of research, but even so sharing data is for varying reasons only slowly becoming the norm. Apart from protecting intellectual property as well as protecting personal information or other sensitive data, a lack of incentives to share could be part of the explanation. In recent years, a number of studies have pointed out the importance of data sharing (Altman & Crosas, 2013; Tenopir et al., 2011, 2015).

The field of Big Data is evolving fast as a fundamental methodology within a number of scientific fields. Prominent examples are Social Science and evidence based medical research (Lynch, 2008). As a result the evaluation of data and elucidation of patterns within the same data become a central part of the research process itself. In order to gain full understanding of the scientific conclusions drawn within the field of Big Data, it is indispensable to gain access to the underlying data. In the field of Computational Science, code sharing is a unique resource that makes it possible to reproduce computerized results as well as allowing the possibility of extending the underlying code. A significant citation advantage was found for articles that shared their code by Vandewalle (2012).

Sharing data requires describing data to make it usable by other researchers, and this is a time consuming process, so if there is no direct mandate, e.g. from a funder, the incentive might have to be quite powerful to convince a researcher to invest the time needed. This incentive could come from a citation advantage and it would be quite powerful—if it exists. Boyce, Biemesderfer, & Owens (1996), Boyce & Biemesderfer (1996) and Boyce (1996) talked about the possibility via online databases for preprints linking to data, and how a “coordinated distributed effort can yield a much more valuable product than any single person or group” (Boyce et al., 1996). But even though citations directly

to research data have increased since 2008, data is still mostly uncited (Peters, Kraker, Lex, Gumpenberger, & Gorraiz, 2016). In the field of astrophysics there still seems to be different opinions about possible citation advantages directly linked to Open Access publications (e.g. Eysenbach, 2006; Kurtz et al., 2005; Kurtz & Henneken, 2007; Swan, 2010). Recently, data citation practices can be studied based on the Data Citation Index by Thomson-Reuters which harvests and registers counts to research data (Robinson-García, Jiménez-Contreras, & Torres-Salinas, 2015). Independent of other factors, citation rates have been shown to increase for studies with publicly available data in case of cancer clinical trials (Piwowar, Day, & Fridsma, 2007) and for studies involving gene expression microarray data (Piwowar & Vision, 2013). Likewise, Henneken and Accomazzi (2011) found higher citation rankings for astronomy papers with links to online data, however they could not differentiate between the different types of papers.

We expand here on a bibliometric study of publications in astrophysics (Dorch, 2012). Dorch looked at data citation advantage over a number of years to papers in a number of astrophysical journals. The aim is to confirm the advantage and to elucidate if there is an inherent citation advantage for types of papers that would typically contain datasets. Part of the present work was presented at The International Astronomical Union's triannual General Assembly in August 2015 and the resulting proceedings (Dorch, Drachen, & Ellegaard, 2015).

The choice of the field of astrophysics for this study has the advantage that it is rather well-defined both in terms of the subject, journals, and institutions – and that astrophysics carries typical traits of other modern Big Science research topics, e.g. displaying both an annual growth in the number of publications and a substantial degree of international collaboration. Furthermore, it can be argued that astrophysics in several ways is both a traditional field of science with roots back to the origin of the scientific method and beyond, but it is also a first-mover, e.g. with respect to adopting Open Science practices including Open Access all the way back to the birth of the Internet.

2. Methods

As described by Dorch (2012) and Dorch et al. (2015), we apply the large, astronomical abstract database Astrophysics Data System (ADS)¹ launched

by the Harvard-Smithsonian Center for Astrophysics in 1992 (Boyce, 1996; Boyce & Biemesderfer, 1996; Boyce et al., 1996; Eichhorn, Kurtz, Accomazzi, Grant, & Murray, 2000; Grant, Accomazzi, Eichhorn, Kurtz, & Murray, 2000; Kurtz et al., 2000). It contains three bibliographic databases with more than 11.8 million records. ADS tracks citations but genuine analyses are not possible at the same level as more traditional citations indexes such as Web of Science (WoS) or Scopus. A unique feature in ADS is the inclusion and registration of links to online data. Search results can then be limited to articles that include these data links.

Sharing of data might be related naturally to publishing the article electronically. Journals published in this fashion have the possibility of being open access as well. In this way, it can be difficult to distinguish if any citation advantage is due to the dataset involved or merely that the paper is published online or open access (Henneken & Accomazzi 2011; Stodden, Guo, & Ma, 2013). In order to avoid this problem, we include only papers from three journals with subscription based access including open access availability after one year. Further, in the present analysis, we compare articles within the same journals, so any difference between online availability can be ruled out.

In the present study we will focus on articles from the three major astronomical journals with a complete coverage in ADS: Astrophysical Journal (ApJ), Astronomical Journal (AJ), and Astronomy and Astrophysics (A&A) as well as the number of citations accumulated during the years 2000–2014. In the ADS database a number of entries such as source titles, keywords, publication years and Digital Object Identifiers (DOIs) of individual articles can be downloaded for further processing in reference management software and citation indexes (cf. Accomazzi & Eichhorn, 2004; Accomazzi, 2011; Eichhorn et al., 2007)

It is necessary to investigate whether we include some ‘bias’ in selecting articles with data-links, i.e. could experimental articles be cited more often than theoretical ones combined with a preference for data-linking by experimental work? Of course, this is only one in a number of possible biases or causal relations for these types of investigations. Other parameters may play a role such as journal type, subject, number of authors, home institution and funding of research. The mere possibility exists that authors preferentially make data available for their best papers (Vandewalle, 2012). In a study by Henneken and Accomazzi (2011), the subject matter was taken into account through a keyword analysis. We choose to consider only whether the article could be

classified as theoretical or as experimental. A full investigation would imply that any two articles with or without d-link must be compared individually with regard to subject.

To test the possibility mentioned above, we entered the DOIs from ADS to the bibliographic database Inspec (by Thomson Reuters) and applied the feature 'treatment type' that this database assigns to all indexed papers:

- "Theoretical or mathematical" (assigned when the subject matter is generally of a theoretical or mathematical nature)
- "Experimental" (used for documents describing an experimental method, observation or result. Includes apparatus for use in experimental work and calculations on experimental results)

In this analysis our focus is only on articles published in 2010, which gives us a citation window of almost 5 years. Articles from the three journals ApJ, A&A and AJ are downloaded into the reference handling program Endnote in order to extract DOIs for further processing. After entering the relevant DOIs into Inspec, the articles are separated into two tiers: Either classified as theoretical or as experimental work. The few articles classified as both experimental and theoretical are discarded from the analysis. Articles classified as 'corrections' or 'editorial' are also not included. These articles gain only none or very few citations and are mainly registered in the segment without data-links. Including the latter articles could skew the analysis but the number actually amounts only to a few percent. Finally, we apply in this case Web of Science (WoS) in order to extract the number of citations because DOIs are not searchable in ADS.

ApJ as registered by ADS includes letters as well as the supplement series but the articles published in those latter categories are not fully included in WoS and we discard them from the present analysis. This discarding is done using Endnote.

We define the citation advantage for articles with datalinks as the ratio of the average number of citations per year to papers with links to data and the average number of citations per year to papers without such links. Alternatively, the citation advantage can be estimated as the ratio between the fraction of citations to articles with links to data and the fraction of articles that have data links (Dorch, 2012). The latter definition gives a smaller number in case of a citation advantage in linking to data.

Statistical analyses were performed as appropriate to test for significant differences in mean citation counts between articles with and without datalinks as well as between theoretical and experimental articles. F-tests were used to test for equal variance and a two sample, two tailed t-test for equal or unequal variance was then used as appropriate.

Further, it is important to discern if any citation advantage could be caused by a very skewed distribution of citations, i.e. is it likely that some few articles with data sharing attract a significantly large number of citations?

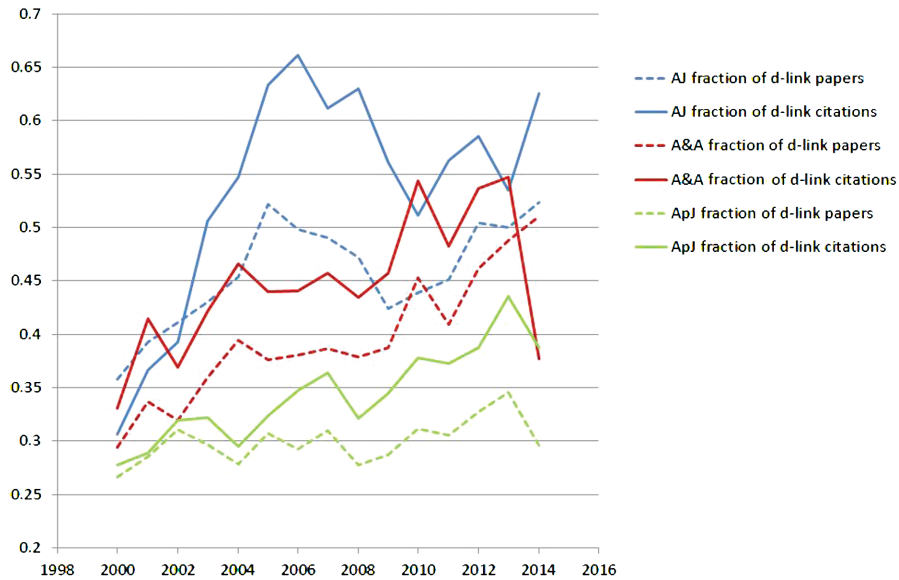
In order to test if a given citation to an article with a data link is generic or actually triggered by the accompanying dataset; seven articles were randomly selected from the 2010 issues of *The Astrophysical Journal*. The in-text references in the citing articles were investigated with respect to the nature of the citation: Where the dataset(s) in the cited paper referred to in the in-text reference in the citing paper? To answer this question the paragraph which included the reference was examined and the cited source was noted. Thus a reference would either score a D-citation or a Non-D-citation. For a reference to score as a D-citation, the data set would have to be mentioned somewhere in the containing paragraph.

Please note that what is counted here is the number of referrals to an article even if this happens multiple times in the text. So these numbers cannot be translated into citations. The weakness of this method is that it could very well be the data themselves that attracted the reference, but the author chose to refer to the article.

3. Results and discussion

For the period 2000–2014, a paper in the set of three journals receives on average 37.1 citations per year. The examined papers in the dataset with datalinks received in total 40.5 citations per year on average, whereas the papers without links to data received correspondingly fewer citations; 35.3 per year. Overall, as presented in Table 1 and Figures 1 and 2, fewer papers link to data and in almost all cases these papers receive more citations per paper than those without data-links, at least during the investigated span of years. Actually, only *Astronomical Journal* has a citation advantage ratio below one and it occurred during the period 2000–2002. Likewise, we have

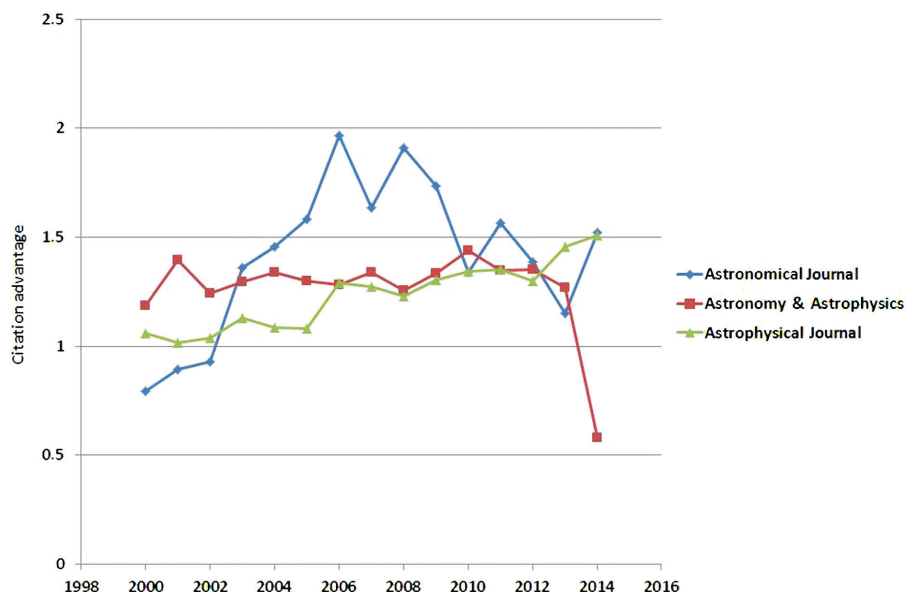
Fig. 1: The fraction of the total number of citations that results from papers with links to data compared to the fraction of papers with data links. Data from ADS for the three journals AJ, A&A and ApJ (including letters and supplements) during the period 2000–2014.



no reasonable explanation of the spurious low citation advantage value ~ 0.6 in case of A&A (Figure 2) in 2014 unless it takes longer times for citations to accumulate in case of data articles although this trend is not evident for AJ and ApJ. The citation advantage of data articles published in ApJ is rising steadily during the period. The similar data for A&A is more constant while the AJ ratio fluctuates during the period. In general, during the period, the data link papers in total received about 25% more citations per paper on average while this number is more like 40% more citations since 2009.

Next, we take a closer look at the three astronomical journals and the articles in terms of links to datasets and their experimental versus theoretical content. Articles with datalinks in ApJ amount to about 31% of all articles published in 2010 in this journal and they are cited more frequently compared to articles without datalinks. Further, the data in Table 2 demonstrates that the number of experimental articles is only well above the number of theoretical ones. The difference between the mean number of citations obtained by the two subsets is rather small as well.

Fig. 2: The citation advantage of papers that links to data as a function of the year of publication as registered in ADS for three different astrophysical journals AJ, A&A and ApJ (including letters and supplements) during the period 2000–2014.



The situation is quite different when we consider articles with or without data links separately. In case of link articles, the number of experimental articles is much larger than the number of theoretical ones, while the latter has the largest number of citations. In contrast, the number of theoretical non-link articles is above the similar number of experimental articles, but, still, the theoretical articles obtain the most citations.

The same pattern is observed in case of the two other journals A&A and AJ (Tables 1 and 3). Both journals have a large number of articles with datalinks, almost 45%, and well above the level for ApJ. The articles with data links obtain the highest number of citations and this trend is further enhanced for theoretical articles. The difference is most pronounced in case of articles published in AJ but this conclusion is based on rather few articles in the dataset.

We summarize the statistical confidence level of our conclusions in Tables 4 and 5. In case of the journals ApJ and A&A, it is proven, well below the 5%

Table 1: Articles published in the year 2010 in *Astrophysical Journal (ApJ)*, *Astronomy and Astrophysics (A&A)* and *Astronomical Journal (AJ)* as registered by ADS.

Journal papers published in 2010	Astro-physical Journal (ApJ)	Astronomy and Astrophysics (A&A)	Astronomical Journal (AJ)
All			
# of papers	2501	1918	388
# of citations	74,663	40,829	9465
Mean citations/paper	29.9	21.3	24.4
Datalink papers			
# of papers	794	875	174
# of citations	27,936	22,308	4754
Mean citations/paper	35.2	25.5	27.3
No datalink papers			
# of papers	1707	1043	214
# of citations	46,727	18,521	4711
Mean citations/paper	27.4	17.8	22.0

Articles are subdivided in this database according to whether they link to data or not.
Citation data from WoS.

significance level ($p < 0.05$) in both cases, that articles with datalinks obtain the largest number of citations. In the case of AJ, the value $p = 0.48$ indicates that the citation advantage is not statistically well founded although articles with datalinks obtain on average more than a 24% higher number of citations compared to articles without datalinks. The large p-value is partly due to the relatively small number of AJ articles included in our sample. In a similar fashion, the p-values indicate a statistically significant advantage in obtaining citations for the theoretical data link articles published in ApJ and AJ ($p = 0.03$ and $p = 0.08$). In case of A&A, theoretical articles with datalinks obtain approximately 15% more citations than experimental articles but the p-value ($= 0.34$) is well above the 5% level. The median values calculated from the citation distributions are also shown in Tables 4 and 5. Our data consistently demonstrates that the citation advantage of data-linked articles also shows up in the calculated median values. This clearly indicates that the effect is not alone caused by a small number of highly cited papers. In contrast, Vandewalle (2012) found that an overwhelming share of the top-cited papers published in two out of three high-profile journals shared their code.

Table 2: Articles published in the year 2010 in *Astrophysical Journal (ApJ)* as registered by ADS.

Astrophysical Journal (ApJ) papers published in 2010	All	Experimental papers	Theoretical papers
All			
# of papers	2501	1312	979
# of citations	74,663	37,300	30,307
Mean citations/paper	29.9	28.4	31.0
Datalink papers			
# of papers	794	614	112
# of citations	27,936	20,143	5005
Mean citations/paper	35.2	32.8	44.7
No datalink papers			
# of papers	1707	698	867
# of citations	46,727	17,157	25,302
Mean citations/paper	27.4	24.6	29.2

Articles are subdivided in this database according to whether they link to data or not. Experimental and theoretical taggings are obtained from Inspec. Citation data from WoS.

The fraction of data datalink papers and fraction of data link citations in case of all three journals are shown in Figure 1. The curves follow the trend already shown in Figure 2. It is evident that articles from AJ in most years constitute the largest fraction of data articles compared to the other journals and the largest citation advantage as well. If we compare A&A and ApJ this tendency is less obvious. The A&A fractions of data link papers are well above the ApJ values but the citation advantage data are similar.

The analysis of seven random articles in the subset of articles with data-links shows that most of the in-text referrals were to either the conclusions of the cited article of a more or less general nature, and only a minority refers directly to the data of the cited article (Figure 3).

The distribution ranges from 0% to 35% of referrals going to data. An article could be referred to because of a conclusion made possible by a compelling data set but without mentioning the data in the referral, and so it would not be counted. But it is possible to speculate that if data citation would be possible, i.e. citations to data sets via a separate DOI different from the article

Table 3: Articles that link to data published in Astrophysical Journal (ApJ), Astronomy and Astrophysics (A&A) and Astronomical Journal (AJ) in 2010 as registered by ADS.

Journal articles with datalink published in 2010	Astro physical Journal (ApJ)	Astronomy and Astrophysics (A&A)	Astronomical Journal (AJ)
All			
# of papers	794	875	174
# of citations	27,936	22,308	4754
Mean citations/paper	35.2	25.5	27.3
Experimental papers			
# of papers	614	720	135
# of citations	20,143	16,312	2737
Mean citations/paper	32.8	22.7	20.4
Theoretical papers			
# of papers	112	95	21
# of citations	5005	2331	1410
Mean citations/paper	44.7	26.1	67.1

Experimental and theoretical taggings are obtained from Inspec. Citation data from WoS.

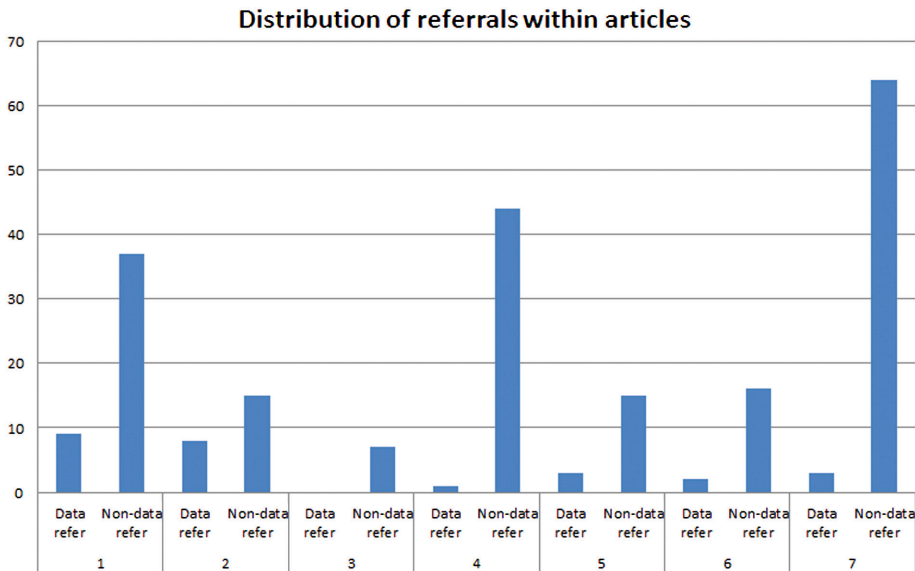
Table 4: Statistics of citations per paper comparing papers with and without data links.

Citations per paper mean (median)	Astrophysical Journal (ApJ)	Astronomy & Astrophysics (A&A)	Astronomical Journal (AJ)
All 2010 articles	29.9 (19)	21.3 (13)	24.4 (11)
All datalink papers	35.2 (22)	25.5 (15)	27.3 (12)
P of T-test	<0.001	<0.001	0.475
No datalink papers	27.4 (18)	17.8 (11)	22.0 (10)

Table 5: Statistics of citations per paper comparing experimental and theoretical papers with data links.

Citations per paper mean (median)	Astrophysical Journal (ApJ)	Astronomy & Astrophysics (A&A)	Astronomical Journal (AJ)
Datalink experimental papers	32.8 (20)	22.7 (15)	20.4 (11)
P of T-test	0.027	0.345	0.072
Datalink theoretical papers	44.7 (28)	26.1 (14)	67.1 (20)

Fig. 3: The distribution of referrals to seven randomly chosen articles to either the data or to other parts of those articles.



DOI, it could be that the citation advantage of the articles themselves would be evened out.

4. Conclusions

We have presented results on the existence of a citation advantage related to sharing astrophysical data. Using bibliographic data from e.g. the NASA ADS database, we find a citation advantage for papers in three representative astrophysical core journals indexed in NASA ADS: The advantage arises when papers are associated with data by datalinks, and consists of datalink papers receiving on the average significantly more citations per paper per year, than papers not associated with links to data. It has not been possible to analyse if papers with datalinks link to collected data or to shared datasets. Further research could look into the reasons for higher citations to datalink papers and specifically if links to own data or collective data has an impact on citation counts.

The citation advantage is evident during the whole period investigated apart from the years 2000–2002 in case of *Astronomical Journal*. Furthermore, we find a tendency for a slightly higher citation advantage for theoretical papers, as opposed to experimental papers. On the other hand, this fact cannot explain the citation advantage of datalink articles. This is due to the fact that the citation advantage is present for both experimental and theoretical articles but the number of theoretical data links are by far the lowest. In the present investigation, we only consider whether the citation advantage was influenced by the work being either experimental or theoretical. Other biases, such as, e.g. a specific field dependence, could influence the outcome as well.

The calculated values of the citation advantage might well be in accordance with the range of referrals going to data as inferred from a minor random number of ApJ data-link articles: It further substantiates the claim that linking to data increases the citation count. Nevertheless, the question remains whether it is the availability of data which leads to citations, rather than the availability of articles? For future studies it could be interesting to investigate if the distribution of referrals to data is different in a sample of articles with no datalinks. However, this would be difficult to carry out since it would be necessary to manually check if there indeed is a data set in those articles. Overall, our citation analysis lends support to the inherent advantage of linking to data in astrophysical research articles.

Acknowledgements

This research has made use of NASA's Astrophysics Data System Bibliographic Services. The authors thank NASA and The University of Southern Denmark.

References

Accomazzi, A., & Eichhorn, G. (2004). Publishing links to astronomical data on-line. In F. Ochsenbein, M.G. Allen, & D. Egret (Eds.), *Astronomical Data Analysis Software and Systems (ADASS) XIII* (pp. 181–184). Proceedings of the conference held 12–15 October, 2003 in Strasbourg, France. ASP Conference Proceedings, Vol. 314. San Francisco: Astronomical Society of the Pacific. Retrieved July 24, 2016, from <http://adsabs.harvard.edu/full/2004ASPC..314..181A>.

Accomazzi, A. (2011). Linking literature and data: Status report and future efforts. In A. Accomazzi (Ed.), *Future professional communication in astronomy II* (pp. 135–142). Astrophysics and Space Science Proceedings, Volume 1. New York: Springer. Retrieved July 24, 2016, from <http://arxiv.org/abs/1103.4295v1>.

Altman, M., & Crosas, M. (2013). The Evolution of data citation: from principles to implementation. *IASSIST Quarterly*, 62(37), 62–70.

Boyce, P. (1996). Journals, data and abstracts make an integrated electronic resource. (189th AAS Meeting, id.06.03). *Bulletin of the American Astronomical Society*, 28, 1280.

Boyce, P.B., & Biemesderfer, C. (1996). Electronic publishing at the American Astronomical Society. In G.H. Jacoby & J. Barnes (Eds.), *Astronomical Data Analysis Software and Systems V*, ASP Conference Services, Vol. 101. Retrieved July 24, 2016 from ftp://iraf.noao.edu/adassconf/web/adass_proc/adass_95/boycep/boycep.html.

Boyce, P.B., Biemesderfer, C., & Owens, E. (1996). The AAS electronic resource. *Vistas in Astronomy*, 40(3), 423–428. [http://dx.doi.org/10.1016/S0083-6656\(96\)00026-8](http://dx.doi.org/10.1016/S0083-6656(96)00026-8).

Dorch, S.B.F. (2012). On the citation advantage of linking to data: astrophysics. Retrieved July 24, 2016, from <https://hal-hprints.archives-ouvertes.fr/hprints-00714715>.

Dorch, S.B.F., Drachen, T.M., & Ellegaard, O. (2015). *The data sharing advantage in astrophysics*. International Astronomical Union. To be published in Conference proceedings of Focus Meeting 3 on Scholarly Publication in Astronomy, IAU GA 2015, Honolulu. Retrieved July 24, 2016, from <http://arxiv.org/abs/1511.02512v1>.

Eichhorn, G., Kurtz, M.J., Accomazzi, A., Grant, C.S., & Murray, S.S. (2000). The NASA astrophysics data system: the search engine and its user interface. *Astronomy and Astrophysics Supplement*, 143(1), 61–83. Retrieved July 24, 2016, from <http://dx.doi.org/10.1051/aas:2000171>.

Eichhorn, G., Accomazzi, A., Grant, C.S., Kurtz, M.J., Thompson, D.M., & Murray, S.S. (2007). Access to the literature and connection to on-line data. *Bulletin of the Astronomical Society of India*, 35, 717–725. Retrieved July 24, 2016, from <http://adsabs.harvard.edu/abs/2007BASl...35..717E>.

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology*, 4(5), e157. Retrieved July 24, 2016, from <http://dx.doi.org/10.1371/journal.pbio.0040157>.

Grant, C.S., Accomazzi, A., Eichhorn, G., Kurtz, M.J., & Murray, S.S. (2000). The NASA astrophysics data system: data holdings. *Astronomy & Astrophysics Supplement Series* 143(1), 111–135. Retrieved July 24, 2016, from <http://dx.doi.org/10.1051/aas:2000173>.

Henneken, E.A., & Accomazzi, A. (2011). Linking to data—effect on citation rates in astronomy. In P. Ballester (Ed.), *Proceedings of ADASS XXI. Book Series: Astronomical*

- Society of the Pacific Conference Series* 461 (pp. 763–766). San Francisco: Astronomical Society of the Pacific. Retrieved July 24, 2016, from <http://arxiv.org/abs/1111.3618>.
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C.S., Murray, S.S., & Watson, J.M. (2000). The NASA astrophysics data system: overview. *Astronomy and Astrophysics Supplement*, 143(1), 41–59. Retrieved July 24, 2016, from <http://dx.doi.org/10.1051/aas:2000170>.
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S.S. (2005). The effect of use and access on citations. *Information Processing & Management*, 41(6), 1395–1402. <http://dx.doi.org/10.1016/j.ipm.2005.03.010>
- Kurtz, M.J., & Henneken, E.A. (2007). Open access does not increase citations for research articles from the astrophysical journal. Retrieved July 24, 2016, from <http://arxiv.org/abs/0709.0896>.
- Lynch, C. (2008). Big data: How do your data grow. *Nature*, 455, 28–29. <http://dx.doi.org/10.1038/455028a>.
- Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. (2016). Research data explored: an extended analysis of citations and altmetrics. *Scientometrics* 170, 723–744. Retrieved July 24, 2016, from <http://dx.doi.org/10.1007/s11192-016-1887-4>.
- Piwowar, H.A., Day, R.S., & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, 2(3), e308. Retrieved July 2016, from <http://dx.doi.org/10.1371/journal.pone.0000308>.
- Piwowar, H.A., & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175; <http://dx.doi.org/10.7717/peerj.175>.
- Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2015). Analyzing data citation practices using the data citation index. *Journal of the Association for Information Science and Technology*. <http://dx.doi.org/10.1002/asi.23529>.
- Stodden, V., Guo, P., & Ma, Z. (2013). Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PLoS One*, 8(6), e67111. Retrieved July 24, 2016, from <http://dx.doi.org/10.1371/journal.pone.0067111>.
- Swan, A. (2010). The open access citation advantage: studies and results to date. Retrieved July 24, 2016, from <http://eprints.soton.ac.uk/id/eprint/268516>.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS One*, 6(6), e21101. Retrieved July 24, 2016, from <http://dx.doi.org/10.1371/journal.pone.0021101>.
- Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in data sharing and data reuse—practices and perceptions among scientists worldwide. *PLoS One*, 10(8), e0134826. Retrieved July 24, 2016, from <http://dx.doi.org/10.1371/journal.pone.0134826>.

Vandewalle, P. (2012). Code sharing is associated with research impact in image processing. *Computing in Science and Engineering*, 14(4), 42–47. <http://dx.doi.org/10.1109/MCSE.2012.63>.

Note

¹ See the ADS Query form (The SAO/NASA Astrophysics Data System), at http://adsabs.harvard.edu/abstract_service.html.